



## Original Research

# Individualized prognosis stratification in muscle invasive bladder cancer: A pairwise *TP53*-derived transcriptome signature

Hua-Ping Liu<sup>a</sup>, Wei Jia<sup>a</sup>, Gaoher Kadeerhan<sup>a</sup>, Bo Xue<sup>a</sup>, Wenmin Guo<sup>a</sup>, Lu Niu<sup>a</sup>, Xiaoliang Wang<sup>a</sup>, Xiaolin Wu<sup>a</sup>, Haitao Li<sup>a</sup>, Jun Tian<sup>a</sup>, Dongwen Wang<sup>a,\*</sup>, Hung-Ming Lai<sup>b,\*</sup>

<sup>a</sup> National Cancer Center, National Clinical Research Center for Cancer, Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen 518116, China

<sup>b</sup> Aiphaqua Genomics Research Unit, Taipei 111, Taiwan

## ARTICLE INFO

## Keywords:

MIBC  
*TP53*  
 13-GPs  
 Prognosis signature  
 Cross-platform

## ABSTRACT

*TP53* is the most frequently mutated gene in muscle invasive bladder cancer (MIBC) and there are two gene signatures regarding *TP53* developed for MIBC prognosis. However, they are limited to immune genes only and unable to be used individually across platforms due to their quantitative manners. We used 827 gene expression profiles from seven MIBC cohorts with varied platforms to build a pairwise *TP53*-derived transcriptome signature, 13 gene pairs (13-GPs). Since the 13-GPs model is a single sample prognostic predictor, it can be applied individually in practice and is applicable to any gene-expression platforms without specific normalization requirements. Survival difference between high-risk and low-risk patients stratified by the 13-GPs test was statistically significant (HR range: 2.26–2.76, all  $P < .0001$ ). Discovery and validation sets showed that the 13-GPs was an independent prognostic factor after adjusting other clinical features (HR range: 2.21–2.82, all  $P < .05$ ). Moreover, it was a potential supplement to the consensus molecular classification of MIBC to further stratify the LumP subtype (patients with better prognoses). High- and low-risk patients by the 13-GPs model presented distinct immune microenvironment and DDR mutation rates, suggesting that it might have the potential for immunotherapy. Being a general approach to other cancer types, this study demonstrated how we integrated gene variants with pairwise gene panels to build a single sample prognostic test in translational oncology.

## Introduction

Bladder cancer is the 10th most common cancer globally, with approximately 573,000 new cases reported in 2020 [1]. Approximately 25% of bladder cancer cases present with muscle-invasive [2,3]. Additionally, up to 50% or more patients with high-risk non-muscle invasive bladder cancer (NMIBC) can progress to MIBC [2,4]. MIBC is a malignancy of high mortality and heterogeneity that needs frequent and long-term surveillance [5–7]. Management of MIBC is still complex and depends on empiricism, so that patients are not routinely stratified and rational therapeutics are exceedingly limited. Consequently, there is a need to identify biomarkers that can aid in stratifying the prognosis of MIBC and thus help assign patients to appropriate treatment approaches.

*TP53* is a tumor suppressor gene located on the short arm of chromosome 17 (17p13.1) and plays a significant role in maintaining genomic stability in response to DNA damage by activating DNA repair

programs and triggering cell-cycle arrest. *TP53* mutations occur in a high percentage of MIBC. The protein product of *TP53*, p53, was reported to be associated with tumor aggressiveness and correlated with poor oncological outcome. Hodgson et al. observed that there is a significant correlation between the contemporary p53 scoring scheme and *TP53* mutations ( $P < .0001$ ) [8]. Esrig et al. showed that overexpression of p53 in the nuclei of tumor cells can provide prognostic information in patients with MIBC [9]. Wang et al. suggested that altered expression of p53 is associated with worse outcome in MIBC [10]. However, MIBC is highly heterogeneous, it is unlikely only use one single p53 marker to adequately predict the prognosis of MIBC. Gene expression, as a bridge between DNA level and protein level, can be applied to identify prognosis molecular markers. Thus, the aim of the present study was to combine *TP53* mutation and gene expression to identify a panel of cross-platform molecular markers for MIBC individualized prognosis stratification.

\* Corresponding authors.

E-mail addresses: [urology2007@126.com](mailto:urology2007@126.com) (D. Wang), [hmlai@aiphaqua.com](mailto:hmlai@aiphaqua.com) (H.-M. Lai).

<https://doi.org/10.1016/j.tranon.2023.101629>

Received 1 December 2022; Received in revised form 3 January 2023; Accepted 16 January 2023

1936-5233/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Table 1**  
Demographic and clinical characteristics for MIBC patients in different cohorts.

Characteristics <sup>a</sup>	TCGA-BLCA	E-MTAB-1803	GSE31684	GSE48075	GSE48276	GSE13507	GSE32894
Number of patients	408	85	78	73	64	62	51
Median survival time <sup>b</sup> (month) (95% CI)	33.1 (26.9 - 54.9)	26.0 (19.0 - 60.0)	58.3 (17.0 - NA)	37.2 (18.7 - 82.4)	47.5 (32.1 - NA)	26.4 (15.1 - NA)	NA (24.2 - NA)
Number of Death (%)	182 (44.6)	43 (50.1)	36 (46.2)	45 (61.6)	34 (53.1)	29 (46.8)	23 (45.1)
Age (Years) <sup>c</sup>	68.1 ± 10.6	68.0 ± 11.1	69.0 ± 10.4	68.8 ± 10.2	66.0 ± 10.3	67.0 ± 9.9	66.0 ± 7.6
Gender <sup>d</sup>							
Female	107	15	21	19 <sup>d</sup>	11	14	13
Male	300	70	57	54 <sup>d</sup>	53	48	38
Grade							
Poorly differentiated	386	80	78	-	-	43	48
Low/moderately	21	5	0	-	-	19	3
T-stage <sup>e</sup>							
T0	1	0	0	0	0	0	0
Ta	-	0	0	0	0	1	0
T1	3	0	0	0	0	0	0
T2	118	31	17	42	13	31	43
T3	194	35	42	23	41	19	7
T4	58	19	19	8	10	11	1
Histology type (%)							
urothelial	408	71	73	-	45	-	-
micropapillary	0	0	0	-	5	-	-
sarcomatoid differentiated	0	0	0	-	1	-	-
squamous differentiated	0	14	5	-	11	-	7
focal glandular	0	0	0	-	2	-	-

<sup>a</sup> Sum of frequency numbers may not equal to the total sample size due to missing or unpredictable values;

<sup>b</sup> Median survival time is incalculable because the mortality at the last follow-up time is less than 50%;

<sup>c</sup> Age is represented as mean ± standard deviation;

<sup>d</sup> Gender of individual patients from cohort of GSE48075 is not offered in the original study;

<sup>e</sup> T stage is given by three forms, respectively, pathological stage, clinical stage and the stage which is not clearly annotated to pathological stage or clinical stage; some patients' T stage in TCGA are not given; The T stage from cohorts of TCGA-BLCA, GSE31684, GSE48276 is given by pathological stage; The T stage from cohorts of GSE48075 is given by clinical stage; The T stage from cohorts of E-MTAB-1803, GSE13507, GSE32894, GSE32548 is not clearly annotated to pathological stage or clinical stage; <sup>f</sup> There are three patients have replicates, each has three replicates, the number of patients: 414 - 3\*2 = 408.

**Table 2**  
TP53 mutation frequency of MIBC in different studies.

	MSK* <sup>‡</sup>	UC	HMS	MSK <sup>†</sup>		TCGA
Cohort Size	109*	81*	19*	70*	24 <sup>#</sup>	411*
TP53 Mutated Cases	62	33	12	48	10	201
TP53 mutation Frequency	56.88%	40.70%	63.20%	68.57%	41.67%	48.91%
Ranking of TP53 mutation	Top 1	Top 1	Top 1	Top 1	Top 1	Top 1

MSK: Memorial Sloan Kettering Cancer Center; UC: University of Chicago; HMS: Harvard Medical School; MSK\*<sup>‡</sup> from [16]; UC from [17]; HMS from [17]; MSK<sup>†</sup> from [18];

\* : primary MIBC tumors;

# : secondary MIBC tumors.

**Table 3**  
Full list of 13-GPs and their coefficients.

Gene Pair <i>l</i>	$g_h$	$g_k$	Coefficients
CCDC77 :: PRUNE	CCDC77	PRUNE	0.53575536
CD93 :: RAB8B	CD93	RAB8B	0.25894147
CDCA7L :: ZNF438	CDCA7L	ZNF438	0.402737482
CNOT6L :: CHD2	CNOT6L	CHD2	-0.268952127
COMP :: GBP4	COMP	GBP4	0.501401855
FZD3 :: HTR7	FZD3	HTR7	-0.390612041
GALNT6 :: GZMA	GALNT6	GZMA	0.248173025
GNPDA1 :: DCBLD2	GNPDA1	DCBLD2	-0.613594279
HIGD1A :: COL4A1	HIGD1A	COL4A1	-0.366063873
KIF23 :: NKG7	KIF23	NKG7	0.264417406
NOL10 :: CXCL13	NOL10	CXCL13	0.479854408
TK1 :: H1FX	TK1	H1FX	0.345474148
TOB1 :: EPS8	TOB1	EPS8	-0.857455054

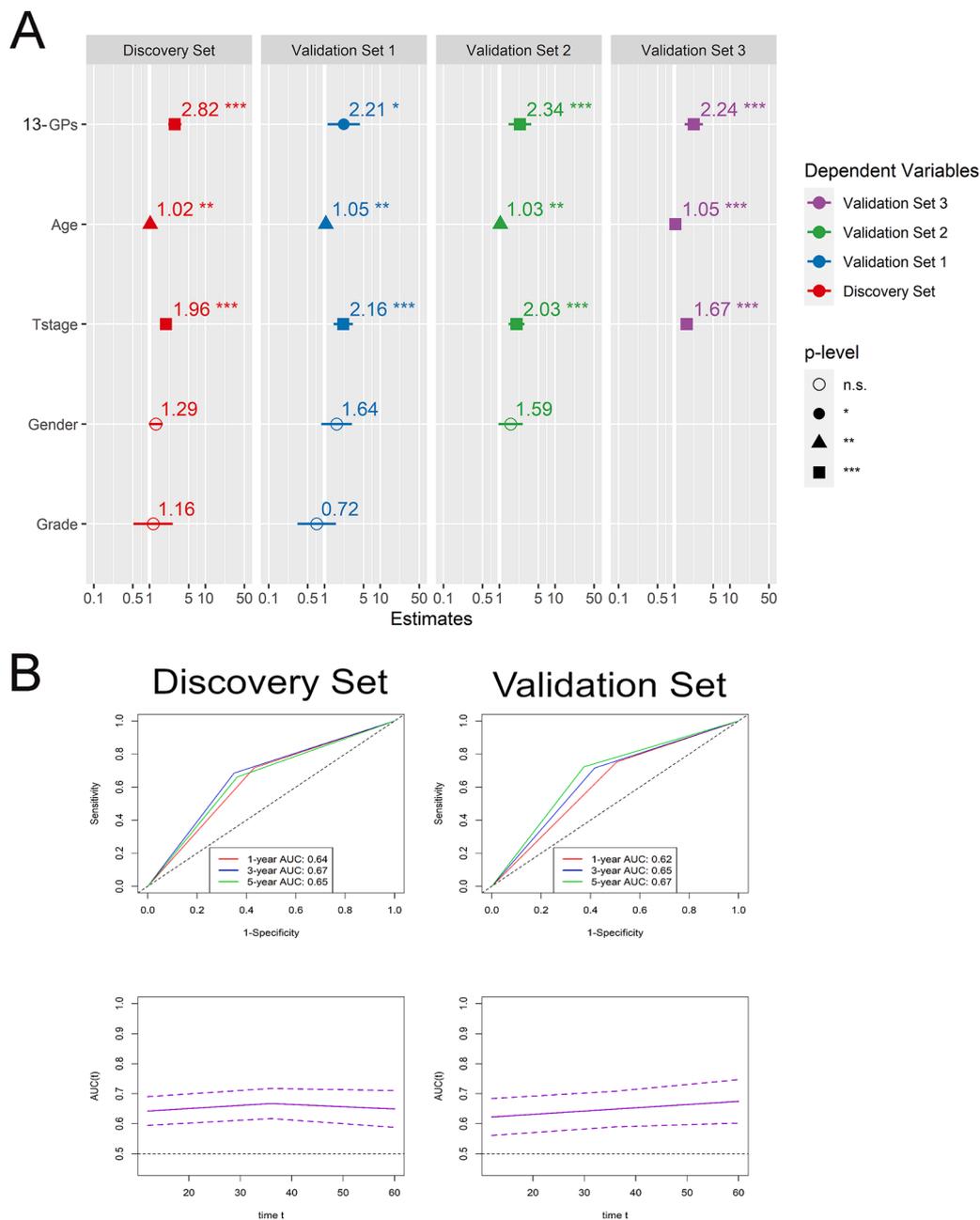
**Patients and methods**

*MIBC datasets collecting and processing*

We collected seven public MIBC gene expression cohorts from different platforms (Table 1, Supplemental Table 1): six microarray cohorts with one from Illumina human-6 v2.0, one from Illumina HumanHT-12 WG-DASL V4.0, two from Affymetrix Human Genome U133 Plus 2.0, and two from Illumina HumanHT-12 V3.0; one RNA-Seq cohort from Illumina HiSeq. Overall survival (OS) was defined as the duration from the date of diagnosis to the date of death from any cause or last follow-up. Only patients with available overall follow-up time, OS status and gene expression data were included. For raw counts of high-throughput sequencing data retrieved from TCGA, Ensembl IDs were transformed to gene symbols and the transcripts per kilobase million (TPM) values were computed. For the microarray data, the probe IDs were annotated to gene symbols according to corresponding platform's annotation file. For multiple probes that mapped to one gene, the mean value of expression was considered.

*Identification and development of 13-GPs*

We used edge R [11] with false discovery rate (FDR) < .05 to select potentially differentially expressed genes (DEGs) between 110 TP53 mutation (TP53MUT) and 304 TP53 wild type (TP53WT) samples upon the TCGA's MIBC cohort. We then built qualitative gene pairs given m (=3775) DEGs  $G = \{g_j\}_{j=1}^m$  and n (=563) samples  $D = \{s_i\}_{i=1}^n$  on the basis of a discovery set (TCGA, E-MTAB-1803, GSE31684). Let  $E = (e_{ij}) \in \mathbb{R}^{n \times m}$  be an expression matrix and  $e_{ij}$  represents an expression value for a gene  $g_j$  on a sample  $s_i$ . For any gene pairs  $l = (g_h, g_k)$ ,  $g_h \neq g_k \in G$ , we defined  $x_{il} = 1$  if  $e_{ih} > e_{ik}$  and  $x_{il} = 0$  otherwise as binary representations on a sample  $s_i$ . A qualitative gene pair  $l$  with  $0 < \sum_{i=1}^n x_{il} < n$  was remained and related to OS for its prognostic assessment by the

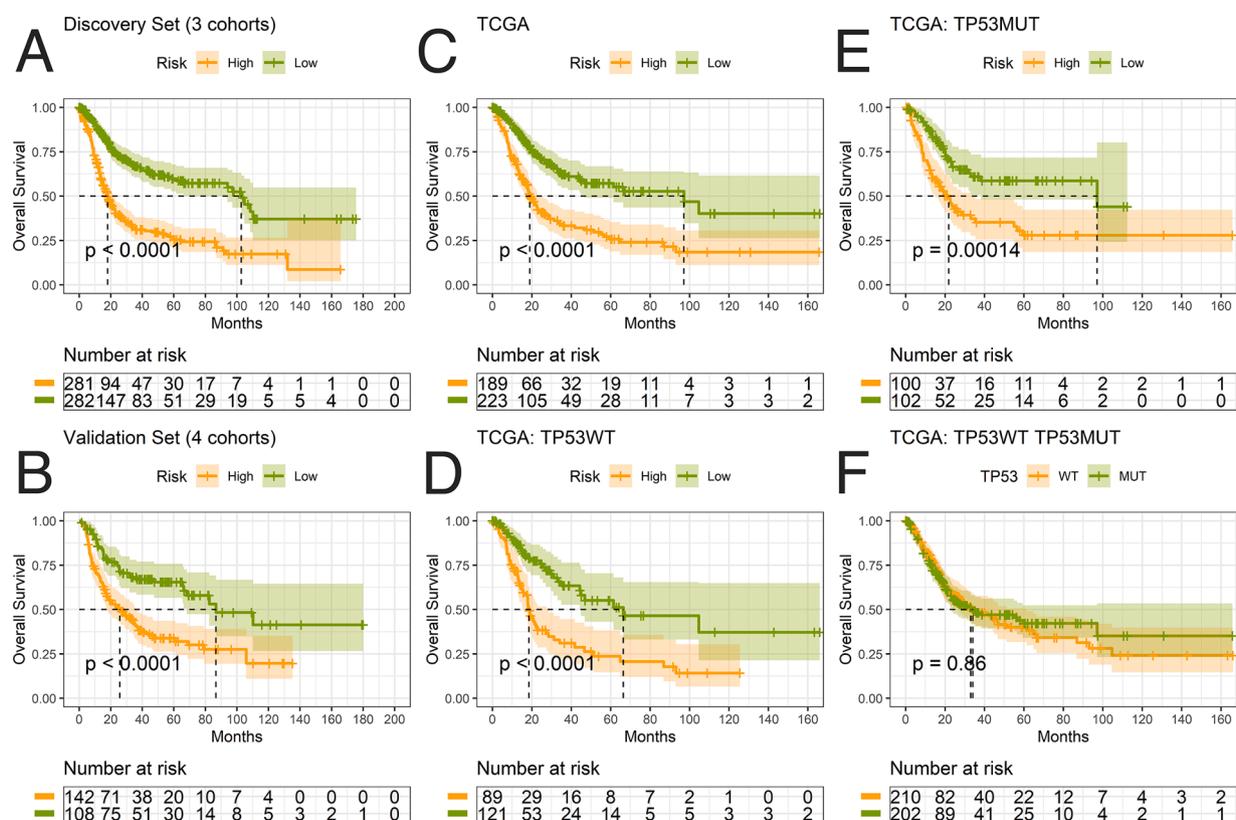


**Fig. 1.** Prognostic value of 13-GPs. (A) Forest plot showing the HR in multivariate Cox proportional hazards regression. Significant p-values are indicated by symbols \*:  $P < .05$ ; \*\*:  $P < .01$ ; \*\*\*:  $P < .001$ . (B) Time-dependent ROC curve analysis at 1-, 3-, and 5-year survival for discovery and validation set.

univariate Cox proportional-hazards model upon the discovery set D. Prognostic gene pairs having  $P$  values less than 0.05 and HR greater than 1.2 or less than 1/1.2 were applied to the multivariate Cox proportional-hazards model with the adaptive least absolute shrinkage and selection operator (adaLASSO) using the “glmnet” R package. Upon the discovery set D, ten-fold cross-validation was conducted to tune a penalty parameter  $\lambda$  and those gene pairs having non-zero coefficients were further examined by the multivariate Cox proportional-hazards model with stepwise selection. Finally, 13 gene pairs and their coefficients were obtained. Given an expression vector  $e^T \in \mathbb{R}^{26}$  of a sample  $s$ , we built a 13-GPs prognostic model to estimate its risk score  $y$  by  $y = \sum_{l=1}^{13} a_l x_l$  where  $a_l$  and  $x_l$  are a coefficient and a binary representation of a gene pair  $l$ , respectively. A cutoff point using a median of the discovery set D was chosen to classify MIBC patients into two categories: low risk ( $y \leq -0.72$ ) and high risk ( $y > -0.72$ ).

**Bioinformatics analysis**

A Kaplan-Meier curve was created for survival analyses using a log-rank test to identify differences in patient survival rates. The OS HR and 95% confidence intervals (CIs) were calculated by Cox regression. Time-dependent ROC curve (timeROC) and Area under timeROC curve (AUC) were generated with R package “timeROC” [12]. CIBERSORT [13] was employed to quantify the immune cell distributions in groups of high-risk and low-risk. The mutation landscape was analyzed by the R package “maftools” following the initial removal of 100 FLAGS genes [13]. DDR genes were acquired from [14]. We used R package “consensusMIBC” to predict individual consensus molecular subtypes for MIBC, including basal/squamous (Ba/Sq), luminal papillary (LumP), luminal unstable (LumU), luminal non-specified (LumNS), neuroendocrine-like (NE-like), and stroma-rich subtypes [15].



**Fig. 2.** OS prognostic analysis in MIBC. (A) Kaplan-Meier survival of the 13-GPs for the discovery set. (B) Kaplan-Meier survival of the 13-GPs for the validation set. (C) Kaplan-Meier survival of the 13-GPs for TCGA cohort. (D) Kaplan-Meier survival of the 13-GPs for TP53WT subgroup in TCGA cohort. (E) Kaplan-Meier survival of the 13-GPs for TP53MUT subgroup in TCGA cohort. (F) Kaplan-Meier OS of *TP53* mutation status.

## Results

### High *TP53* mutation frequency in MIBC

We reviewed *TP53* mutation frequency from five published MIBC cohorts, respectively, MSK\* [16], UC [17], HMS [17], MSK† [18] and TCGA (Table 2). Table 2 shows that no matter the cohort size is small or large, *TP53* is the most commonly mutated gene in all of the five published MIBC cohorts, presenting in 40% - 68% of patients. Among the five cohorts, the size of the TCGA cohort was the largest, and the information of MIBC genome was detailed, we thus further analyzed MIBC genome characteristics in TCGA cohort. DNA mutations of 394 MIBC patients in TCGA database, which are from MC3 were analyzed by maftools package [19] in this study. On gene-level analysis, the top 10 mutated genes included *TP53*, *KMT2D*, *KDM6A*, *ARID1A*, *PIK3CA*, *KMT2C*, *RB1*, *RYR2*, *EP300* and *FAT4*. In detail, mutations in the *TP53* were the most commonly acquired mutations (49%). And the rest of genes mutation frequency sharply decreased under 30% (Supplementary Fig. 1A). Given the possibility that co-occurring or exclusively mutation might contribute together to tumor formation and progression, we next explored the somatic interactions between *TP53* and the remaining top 24 high frequency mutation genes in MIBC (Supplementary Fig. 1B). The results demonstrated that *TP53* was mutually exclusive with *FGFR3* ( $P < .01$ ) but co-occurred with *KMT2A* ( $P < .05$ ), *BIRC6* ( $P < .01$ ), *FAT4* ( $P < .01$ ), and *RB1* ( $P < .01$ ). *FGFR3* is one of the most commonly mutated genes in MIBC and its mutants are significantly correlated with lower-grade bladder cancer with favorable prognosis [20], while *RB1* mutations which frequently co-occur with *TP53* mutations have adverse prognostic significance in advanced bladder cancer [21]. High *TP53* mutation frequency and the above somatic interaction analysis suggested that *TP53* mutation might be a possible prognostic factor in MIBC.

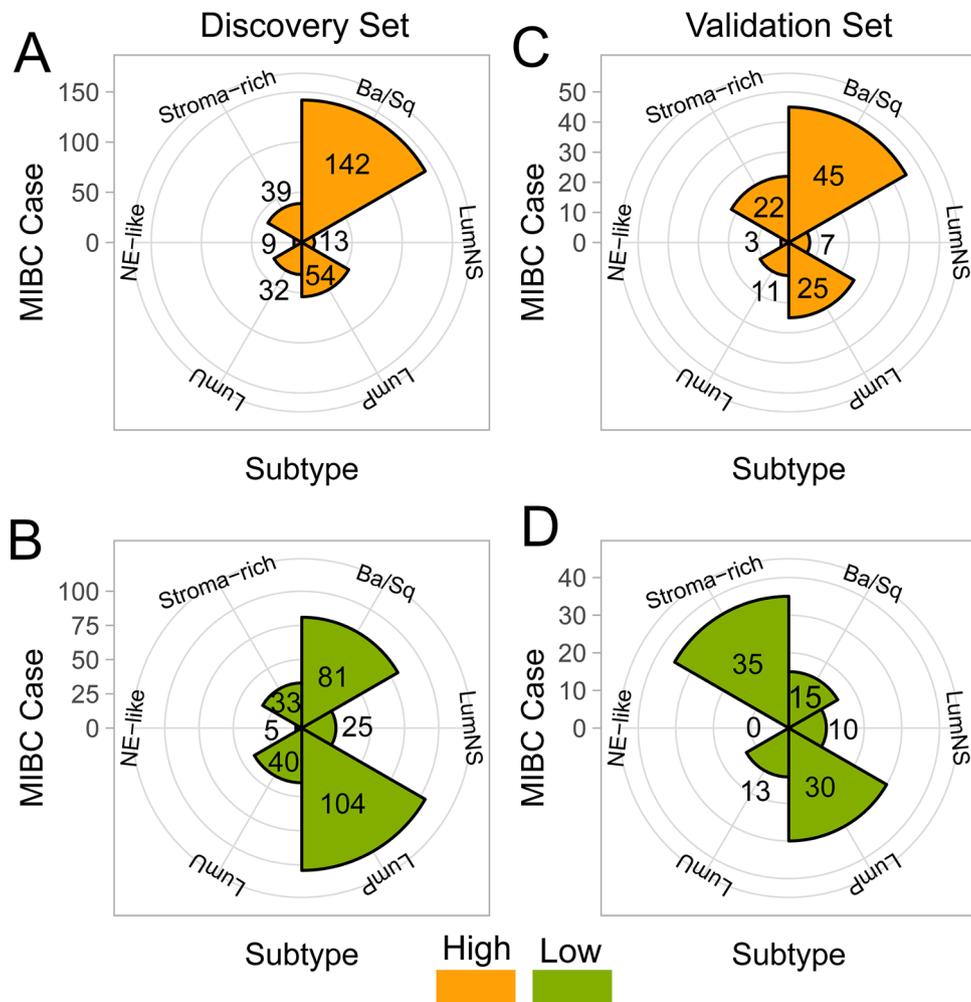
### A pairwise *TP53*-derived gene panel predicts prognosis in patients with MIBC

The 13-GPs prognostic model included 4 immune genes (*GZMA*, *CXCL13*, *COMP*, *NKG7*) and 22 non-immune genes. The full list of 13 gene pairs,  $l = (g_h, g_k)$ , and their coefficients were summarized in Table 3. Univariate and multivariate Cox regression analyses showed that the 13-GPs was an independent prognosis factor for patients with MIBC (Fig. 1A, supplementary Table 2, supplementary Table 3). AUC values estimated by time-dependent ROC curve at 1-, 3-, and 5-year OS in the discovery (AUC = 0.64, 0.67 and 0.65; Fig. 1B) and validation (AUC = 0.62, 0.65 and 0.67; Fig. 1B) sets also indicated that 13-GPs held promising accuracy for OS prediction in MIBC individualized prognosis stratification.

We eventually showed how the pairwise panel indicated the prognostic stratification of MIBC. In the discovery set, Cumulative Kaplan-Meier curves showed that high-risk prognostic group with the higher risk score had significantly poorer OS prognosis than low-risk prognostic group with the low-risk score ( $P < .0001$ ) (Fig. 2A), we also further observed the panel prediction ability in cohorts of TCGA, TP53WT and TP53MUT (Fig. 2C-E). To further evaluate the robustness of 13-GPs, a validation set which consisted of four cohorts was also assessed. The patients in the validation set were stratified into prognostic groups of the low-risk and the high-risk, while the high-risk prognostic group showed a significantly poorer OS survival than the low-risk group ( $P < .0001$ ) (Fig. 2B). Overall, the 13-GPs could estimate OS independently of clinical factors in MIBC.

*13-GPs was not only an effective transcriptome signature for MIBC prognosis stratification but also a supplement for consensusMIBC on Lump*

Cohort-specific high-risk groups were more likely to be subtype of



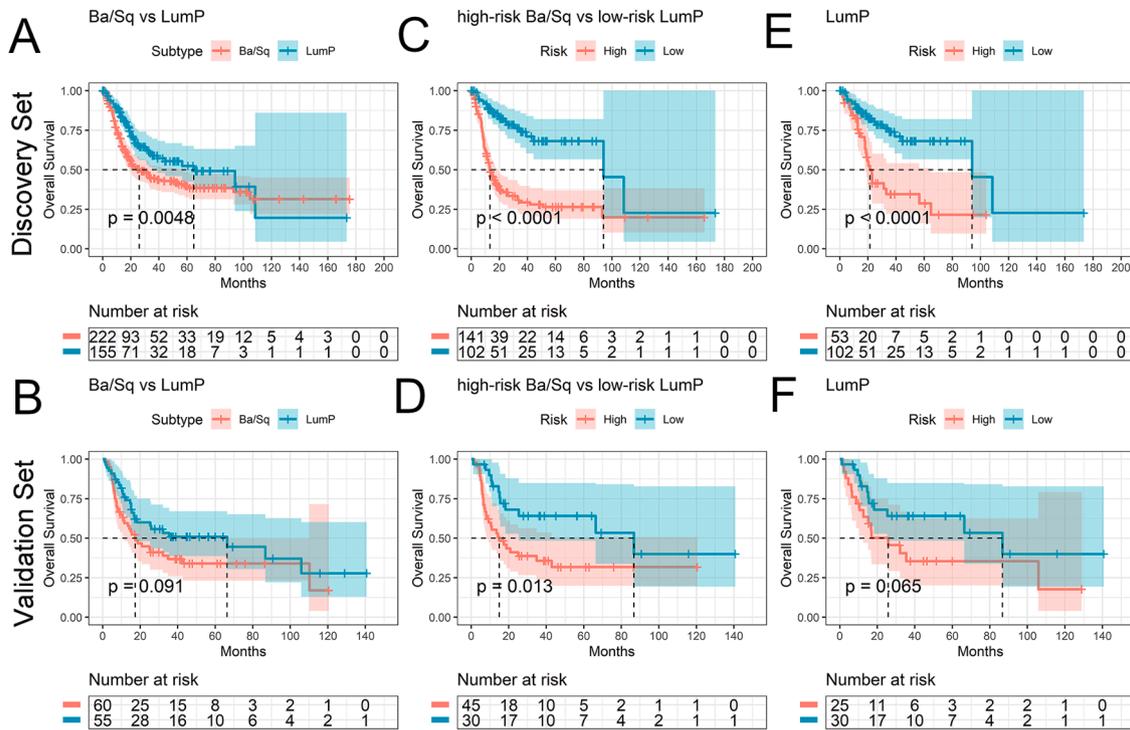
**Fig. 3.** Consensus subtype distribution in groups of high-risk and low-risk MIBC patients. (A) Consensus subtype distribution of high-risk patients in discovery set. (B) Consensus subtype distribution of low-risk patients in discovery set. (C) Consensus subtype distribution of high-risk patients in validation set. (D) Consensus subtype distribution of high-risk patients in validation set.

Ba/Sq with poor median OS, whereas most samples within low-risk groups were predicted as subtype of LumP with relatively good median OS (Fig. 3). Kamoun et al. confirmed OS is strongly associated with consensus subtypes [15]. In our study, Ba/Sq with poor overall prognosis accounted for the highest proportion, followed by LumP with relatively good prognosis. Significantly differential OS between all Ba/Sq and all LumP in our study was observed (Fig. 4A and B,  $P = .0048$  in discovery set,  $P = .091$  in validation set). Differential OS between high-risk Ba/Sq and low-risk LumP in our study was more significant (Fig. 4C and D,  $P < .0001$  in discovery set,  $P = .013$  in validation set). Interestingly, OS of LumP between subgroup of the high-risk and the low-risk was also significantly differential (Fig. 4E and F,  $P < .0001$  in discovery set,  $P = .065$  in validation set) which indicated LumP was heterogeneous but 13-GPs could be a supplement for consensusMIBC to evaluate LumP prognosis.

#### High-risk and low-risk patients with MIBC showed distinct immune microenvironment and DDR mutation rates

Immunotherapy holds promise for treatment of MIBC [22]. The immune cell proportions and immune checkpoints (ICs) characterizations in MIBC are critical for prediction of treatment responses. Using the CIBERSORT in combination with the LM22 signature matrix, we estimated the differences in the immune infiltration of 22 immune cell types between groups of high-risk and low-risk MIBC patients, respectively in

RNA-seq and array data. In the RNA-seq set, the high-risk MIBC patients had significantly higher proportions of macrophages M0 ( $P < .0001$ ), macrophages M2 ( $P < .0001$ ), mast cells activated ( $P = .015$ ), neutrophils ( $P = .0052$ ), T cells CD4 memory resting ( $P = .0038$ ), and significantly lower proportions of NK cells activated ( $P = .0073$ ), T cells CD4 memory activated ( $P = .036$ ), T cells CD8 ( $P < .0001$ ), T cells follicular helper ( $P = .0011$ ), T cells gamma delta ( $P = .0094$ ), and T cells regulatory (Tregs) ( $P = .0093$ ) than the low-risk MIBC patients. In array set, the high-risk MIBC patients had significantly higher proportions of dendritic cells resting ( $P = .015$ ), macrophages M0 ( $P = .01$ ), and neutrophils ( $P = .033$ ), and significantly lower proportions of plasma Cells ( $P = .0091$ ), T cells CD4 memory activated ( $P = .0021$ ), T cells CD8 ( $P = .0073$ ), and T cells gamma delta ( $P < .0001$ ) than the low-risk MIBC patients. The high-risk MIBC patients both in RNA-seq and array data had significantly higher proportions of neutrophils and macrophages M0, and significantly lower proportions of T cells CD8, T cells CD4 memory activated and T cells gamma delta than the low-risk MIBC patients (Fig. 5A). Further, we investigated ICs expression between the prognostic groups of the high-risk and the low-risk in TCGA. Both stimulatory checkpoint ( $CD27$ ,  $CD40$ ,  $ICOS$ ) and inhibitory checkpoint molecules ( $CD276$ ,  $BTLA$ ,  $CTLA4$ ,  $PDCD1$ ,  $SIGLEC9$ ) were differentially expressed which might be served as potential immunotherapy targets for MIBC (Fig. 5B). Compared to the low-risk prognostic MIBC patients, all the checkpoints were up regulated except  $CD276$  and  $SIGLEC9$  in the high-risk prognostic MIBC patients. Results of the immune cell



**Fig. 4.** OS prognostic analysis in Consensus subtype. (A) Kaplan-Meier survival of 13-GPs between Ba/Sq and LumP in discovery set. (B) Kaplan-Meier survival of 13-GPs between Ba/Sq and LumP in validation set. (C) Kaplan-Meier survival of 13-GPs between high-risk Ba/Sq and low-risk LumP in discovery set. (D) Kaplan-Meier survival of 13-GPs between high-risk Ba/Sq and low-risk LumP in validation set. (E) Kaplan-Meier survival of 13-GPs between high-risk LumP and low-risk LumP in discovery set. (F) Kaplan-Meier OS between high-risk LumP and low-risk LumP in validation set.

proportions and ICs indicated that 13-GPs could guide MIBC immunotherapy. DDR and the immune system are tightly interrelated [23,24]. We further observed the DDR mutation landscape respectively in the high-risk and the low-risk patients with MIBC. DDR mutations were more common in low-risk group with an improved OS compared to high-risk group (Fig. 5C). DDR mutation characteristics in groups of high-risk and low-risk suggested that alterations in DNA damage were associated with improved clinical outcomes.

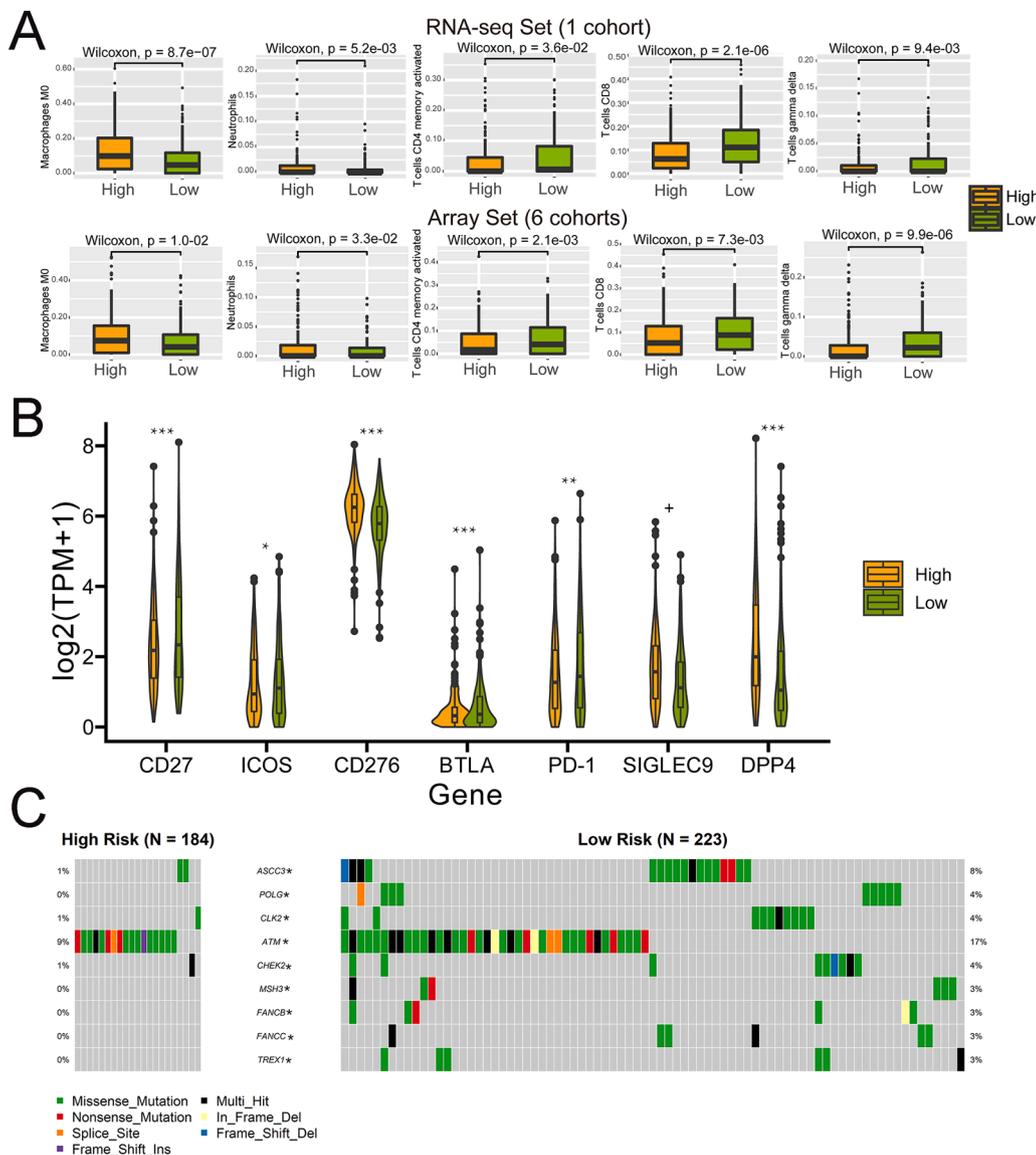
**Discussion**

MIBC is primarily characterized by high heterogeneity. Prognosis stratification in MIBC is necessary so that patients can receive individualized treatment. Clinically, TP53 mutations have been linked to a poorer prognosis for some kinds of cancers [25,26]. And TP53 is the most commonly mutated genes in MIBC [16–18,27]. In order to determine the prognosis of TP53 mutation status in MIBC, we conducted a stratification analysis between TP53WT and TP53MUT. The Kaplan-Meier survival analysis showed that their OS was not significantly differential (Fig. 2F), which was consistent with a result of Donehower et al. [28]. However, p53, a TP53-derived protein, was validated as a prognostic factor in MIBC [8–10]. The p53 tumor suppressor protein can work as a key failsafe mechanism of cellular anti-cancer defenses that inhibits cell division or survival in response to various stresses [29,30]. TP53 mutation, as well as overexpression of key p53 regulatory proteins such as MDM2 can inactivate the p53 protein function [31,32]. Thus, developing an expression signature derived from TP53 mutation that might be more prognostically predictive. A few studies tried to identify a TP53-derived transcriptome signature to stratify MIBC [33,34]. However, it is hard to use them in real clinical practices and their stratification performance based on genes' quantitative expression need to be improved.

In recent years, pairwise gene panels are widely explored in tumor diagnosis and prognosis which have natural advantages in clinical

practice [35,36]: (i) they can be used in different platforms; (ii) they are robust against experimental and technical variations; (iii) they can be applied at an individual level. In the present study, we took advantage of GPs which were used in our previous study [35,36] to construct a pairwise TP53-derived transcriptome signature for MIBC stratification. Although the prognostic stratification ability of 13-GPs did not perform as well as the diagnostic usage in [35,36], it effectively stratified the MIBC patients as groups of the high-risk and the low-risk and further showed distinct immune microenvironment and DDR mutation rates.

Histopathologically, 634 patients were pure urothelial carcinoma, 37 patients had squamous differentiation variants, 5 patients were urothelial carcinoma with micropapillary features, 1 patient had sarcomatoid differentiation, and 2 patients had a focal glandular pattern. We observed that 29 of 37 squamous differentiation variants were classified as high-risk patients, implying that urothelial carcinomas with squamous differentiation were more aggressive. The result was also supported by Tripathi's study [37]. At the suggestion of an anonymous reviewer, we were referred to Zheng et al. [38] for discussing how it is different from our work. Cancer-immune phenotypes are assumed to be cold (immune-desert), immune-excluded or hot (inflamed) [39]. A recent study on 258 MIBC patients shows that cold tumors are in the majority (63%) while 21% and 16% of phenotypes are immune-excluded and hot tumors, respectively [40]. The phenotypes are presented by cancer cells interacting with immune cells and we hypothesized that the interactions should be reflected in synergies between immune and non-immune genes. It would be questionable that Zheng et al. [38] merely use immune genes for a prognostic model to explain phenotypic heterogeneity in MIBC. Our 13-GPs was derived from whole genome scale and included 4 immune genes and 22 non-immune genes while Zheng's model employs 74 immune genes to form 45 gene pairs. This provided experimental evidence indicating that more genes are accumulated in order to explain cancer-immune cell interactions by just immune genes than by synergies between immune and non-immune genes. Certainly, our model having fewer genes is



**Fig. 5.** Immune microenvironment and DDR mutation rates in high-risk and low-risk MIBC patients. (A) Significantly differential proportion of immune infiltration between high-risk and low-risk MIBC patients both in RNA-seq and microarray set. (B) Box plots visualizing significantly different immune checkpoint between high-risk and low-risk patients. <sup>+</sup>FDR < .1, \*FDR < .05, \*\*FDR < .01, \*\*\*FDR < .001 (C) Significantly DDR alterations between high-risk and low-risk MIBC patients, \*FDR < .05.

more practical and more economical in terms of translational implications. Our work had the following limitations. The information regarding patient treatment and histopathology was incomplete, and the association among patient treatment, histopathology and the signature should extensively be investigated. Collection of clinical data in our study was retrospective. Prospective studies are required to validate our model for use in clinical setting.

**Conclusions**

In conclusion, 13-GPs was a simple and useful signature for MIBC individualized prognosis stratification. Not only was 13-GPs adaptable to many platforms, but it also complemented consensusMIBC to improve prognostic risk stratification in LumP and support personalized medicine.

**Funding**

Project funded by China Postdoctoral Science Foundation (No. 2022M713274), Shenzhen High-level Hospital Construction Fund and Sanming Project of Medicine in Shenzhen (No. SZSM202111003).

**Ethics approval**

Not applicable.

**Data sharing**

The datasets supporting the conclusions of this article are publicly available in the TCGA, ArrayExpress (E-MTAB-1803) and Gene Expression Omnibus (GEO) databases (GSE31684, GSE48075, GSE48276, GSE13507 and GSE32894).

**CRediT authorship contribution statement**

**Hua-Ping Liu:** Conceptualization, Methodology, Investigation, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Wei Jia:** Writing – review & editing. **Gaohaer Kadeerhan:** Writing – review & editing. **Bo Xue:** Writing – review & editing. **Wenmin Guo:** Writing – review & editing. **Lu Niu:** Writing – review & editing. **Xiaoliang Wang:** Writing – review & editing. **Xiaolin Wu:** Writing – review & editing. **Haitao Li:** Writing – review & editing. **Jun Tian:** Writing – review & editing. **Dongwen Wang:** Conceptualization, Investigation, Writing – review & editing,

Supervision. **Hung-Ming Lai:** Conceptualization, Methodology, Investigation, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

#### Declaration of Competing Interest

The authors declare that there are no conflicts of interest.

#### Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.tranon.2023.101629](https://doi.org/10.1016/j.tranon.2023.101629).

#### References

- [1] H. Sung, J. Ferlay, R.L. Siegel, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA Cancer J. Clin.* 71 (2021) 209–249.
- [2] S.S. Chang, B.H. Bochner, R. Chou, et al., Treatment of non-metastatic muscle-invasive bladder cancer: AUA/ASCO/ASTRO/SUO guideline, *J. Urol.* 198 (2017) 552–559.
- [3] M.E. Nielsen, A.B. Smith, A.M. Meyer, et al., Trends in stage-specific incidence rates for urothelial carcinoma of the bladder in the United States: 1988 to 2006, *Cancer* 120 (2014) 86–95.
- [4] B.W. van Rhijn, M. Burger, Y. Lotan, et al., Recurrence and progression of disease in non-muscle-invasive bladder cancer: from epidemiology to treatment strategy, *Eur. Urol.* 56 (2009) 430–442.
- [5] M.A. Knowles, C.D. Hurst, Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity, *Nat. Rev. Cancer* 15 (2015) 25–41.
- [6] H.W. Kang, W.J. Kim, W. Choi, S.J. Yun, Tumor heterogeneity in muscle-invasive bladder cancer, *Transl. Androl. Urol.* 9 (2020) 2866–2880.
- [7] W. Choi, S. Porten, S. Kim, et al., Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy, *Cancer Cell* 25 (2014) 152–165.
- [8] A. Hodgson, B.W.G. Van Rhijn, S.S. Kim, et al., Reassessment of p53 immunohistochemistry thresholds in invasive high grade bladder cancer shows a better correlation with TP53 and FGFR3 mutations, *Pathol. Res. Pract.* 216 (2020), 153186.
- [9] D. Esrig, D. Elmajian, S. Groshen, et al., Accumulation of nuclear p53 and tumor progression in bladder cancer, *N. Engl. J. Med.* 331 (1994) 1259–1264.
- [10] G. Wang, P.C. Black, P.J. Goebell, et al., Prognostic markers in pT3 bladder cancer: a study from the international bladder cancer tissue microarray project, *Urol. Oncol.* 39 (2021) 301, e317–301 e328.
- [11] M.D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data, *Genome Biol.* 11 (2010) R25.
- [12] P. Blanche, J.F. Dartigues, H. Jacqmin-Gadda, Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks, *Stat. Med.* 32 (2013) 5381–5397.
- [13] B. Chen, M.S. Khodadoust, C.L. Liu, et al., Profiling Tumor infiltrating immune cells with cibersort, *Methods Mol. Biol.* 1711 (2018) 243–259.
- [14] T.A. Knijnenburg, L. Wang, M.T. Zimmermann, et al., Genomic and molecular landscape of DNA damage repair deficiency across the cancer genome atlas, *Cell Rep.* 23 (2018) 239–254, e236.
- [15] A. Kamoun, A. De Reynies, Y. Allory, et al., A consensus molecular classification of muscle-invasive bladder cancer, *Eur. Urol.* 77 (2020) 420–433.
- [16] P.H. Kim, E.K. Cha, J.P. Sfakianos, et al., Genomic predictors of survival in patients with high-grade urothelial carcinoma of the bladder, *Eur. Urol.* 67 (2015) 198–201.
- [17] K.L. Yap, K. Kiyotani, K. Tamura, et al., Whole-exome sequencing of muscle-invasive bladder cancer identifies recurrent mutations of UNC5C and prognostic importance of DNA repair gene mutations on survival, *Clin. Cancer Res.* 20 (2014) 6605–6617.
- [18] E.J. Pietzak, E.C. Zabor, A. Bagrodia, et al., Genomic differences between “primary” and “secondary” muscle-invasive bladder cancer as a basis for disparate outcomes to cisplatin-based neoadjuvant chemotherapy, *Eur. Urol.* 75 (2019) 231–239.
- [19] A. Mayakonda, D.C. Lin, Y. Assenov, et al., Maftools: efficient and comprehensive analysis of somatic variants in cancer, *Genome Res.* 28 (2018) 1747–1756.
- [20] B.W. van Rhijn, A.N. Vis, T.H. van der Kwast, et al., Molecular grading of urothelial cell carcinoma with fibroblast growth factor receptor 3 and MIB-1 is superior to pathologic grade for the prediction of clinical outcome, *J. Clin. Oncol.* 21 (2003) 1912–1921.
- [21] M. Gallucci, F. Guadagni, R. Marzano, et al., Status of the p53, p16, RB1, and HER-2 genes and chromosomes 3, 7, 9, and 17 in advanced bladder cancer: correlation with adjacent mucosa and pathological parameters, *J. Clin. Pathol.* 58 (2005) 367–371.
- [22] J. Kaur, W. Choi, D.M. Geynisman, et al., Role of immunotherapy in localized muscle invasive urothelial cancer, *Ther. Adv. Med. Oncol.* 13 (2021), 17588359211045858.
- [23] K.W. Mouw, M.S. Goldberg, P.A. Konstantinopoulos, A.D. D’Andrea, DNA damage and repair biomarkers of immunotherapy response, *Cancer Discov.* 7 (2017) 675–693.
- [24] G. Chatzinikolaou, I. Karakasilioti, G.A. Garinis, DNA damage and innate immunity: links and trade-offs, *Trends Immunol.* 35 (2014) 429–435.
- [25] A.I. Robles, C.C. Harris, Clinical outcomes and correlates of TP53 mutations and cancer, *Cold Spring Harb. Perspect. Biol.* 2 (2010), a001016.
- [26] F. Meric-Bernstam, X. Zheng, M. Shariati, et al., Survival outcomes by TP53 mutation status in metastatic breast cancer, *JCO Precis. Oncol.* (2018) 2018.
- [27] M. Mossanen, F.L.F. Carvalho, V. Muralidhar, et al., Genomic features of muscle-invasive bladder cancer arising after prostate radiotherapy, *Eur. Urol.* 81 (2022) 466–473.
- [28] L.A. Donehower, T. Soussi, A. Korkut, et al., Integrated analysis of TP53 gene and pathway alterations in the cancer genome atlas, *Cell Rep.* 28 (2019) 3010.
- [29] K.H. Vousden, C. Prives, Blinded by the light: the growing complexity of p53, *Cell* 137 (2009) 413–431.
- [30] E.R. Kasthuber, S.W. Lowe, Putting p53 in context, *Cell* 170 (2017) 1062–1078.
- [31] K.D. Sullivan, M.D. Galbraith, Z. Andrysiak, J.M. Espinosa, Mechanisms of transcriptional regulation by p53, *Cell Death Differ.* 25 (2018) 133–143.
- [32] L.A. Donehower, T. Soussi, A. Korkut, et al., Integrated analysis of TP53 gene and pathway alterations in the cancer genome atlas, *Cell Rep.* 28 (2019) 1370–1384, e1375.
- [33] H. Li, H. Lu, W. Cui, et al., A TP53-based immune prognostic model for muscle-invasive bladder cancer, *Aging* 13 (2020) 1929–1946 (Albany NY).
- [34] X. Wu, D. Lv, C. Cai, et al., A TP53-associated immune prognostic signature for the prediction of overall survival and therapeutic responses in muscle-invasive bladder cancer, *Front. Immunol.* 11 (2020), 590618.
- [35] H.P. Liu, H.M. Lai, Z. Guo, Prostate cancer early diagnosis: circulating microRNA pairs potentially beyond single microRNAs upon 1231 serum samples, *Brief. Bioinform.* 22 (2021), bbaa111.
- [36] H.P. Liu, D. Wang, H.M. Lai, Can we infer tumor presence of single cell transcriptomes and their tumor of origin from bulk transcriptomes by machine learning? *Comput. Struct. Biotechnol. J.* 20 (2022) 2672–2679.
- [37] N. Tripathi, Y. Jo, A. Tripathi, et al., Genomic landscape of locally advanced or metastatic urothelial carcinoma with squamous differentiation compared to pure urothelial carcinoma, *Urol. Oncol.* 40 (2022) 493, e491–493 e497.
- [38] X. Zheng, X. Zhou, H. Xu, et al., A novel immune-gene pair signature revealing the tumor microenvironment features and immunotherapy prognosis of muscle-invasive bladder cancer, *Front. Genet.* 12 (2021), 764184.
- [39] D.S. Chen, I. Mellman, Elements of cancer immunity and the cancer-immune set point, *Nature* 541 (2017) 321–330.
- [40] Y. Zhu, H. Fu, Z. Liu, et al., Immune-desert, immune-excluded and inflamed phenotypes predict survival and adjuvant chemotherapy response in patients with MIBC, *Eur. Urol. Suppl.* 17 (2018).